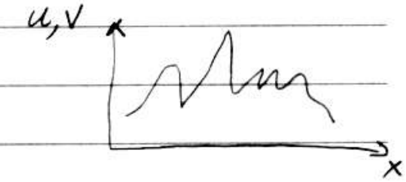
Tutorial 1 : What is an adjoint, why is it useful?

Green's Identity



Continuous space:

inner-product: $\int v u \, dx = \{v, u\}$

which has an associated norm, L2-norm or Eucledean norm

Adjoint operator and the Green's identity:

$$\int v (Au) \, dx = \int u (A^{+} v) \, dx$$

or $\{v, Au\} = \{u, A^{+} v\}$

$A =$ is the operator
$A^{+} =$ is the adjoint of A

The adjoint operator depend on the definition of the inner product.

General Case: Operator M, differt norm

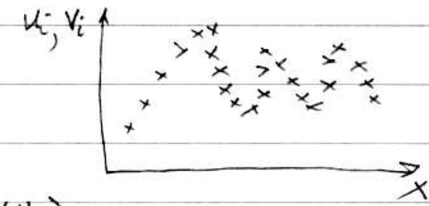
$$\int v M u \, dx = (v, u)$$

Green's identity: $\int v M (Au) \, dx = \int u M (\tilde{A} v) \, dx$

$$\tilde{A} = M^{-1} A^{+} M$$

Discrete space: vectors and matrices (instead of functionals and operators)

$$\int v u \, dx \rightarrow \lim_{\Delta x \to 0} \sum_{i=1}^{N} v_i u_i \Delta x$$



$$\rightarrow \underline{v}^T \underline{u}$$

$$(v_1, v_2, v_3, \ldots) \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \end{pmatrix}$$

(note the absence of the Δx terms) usually ignored, but critically important in some cases to insure $\{v, u\}$ remains finite.

Nomenclature:

vector, v : $\underline{v}$

operator, A : $\underline{\underline{A}}$ matrices

Green's identity

$$\int v A u \, dx = \{v, Au\}$$

$$\rightarrow \underline{v}^T \underline{\underline{A}} \, \underline{u}$$

For every operator there is a discrete analog:

$$\frac{d}{dx} \Rightarrow \frac{1}{\Delta x} \begin{pmatrix} 0 & -1 & & & 0 \\ 1 & 0 & -1 & & \\ & 1 & 0 & -1 & \\ & & 1 & 0 & -1 \\ 0 & & & 1 & \end{pmatrix}$$

$$\underline{v}^T (\underline{\underline{A}} \, \underline{u}) = (\underline{\underline{A}} \, \underline{u})^T \underline{v} = \underline{u}^T \underline{\underline{A}}^T \underline{v}$$

For the L-2 norm $A^+ \equiv \underline{\underline{A}}^T$ is the transpose operator

For the General case $\int v M (Au) \, dx = (v, Au)$

$$\Rightarrow \underline{v}^T [\underline{\underline{M}} (\underline{\underline{A}} \, \underline{u})] = \underline{u}^T \underline{\underline{A}}^T \underline{\underline{M}}^T v$$

$$= u^T \underline{\underline{M}} \underbrace{(\underline{\underline{M}}^{-1} A^T \underline{\underline{M}}^T \underline{v})}_{\tilde{A}}$$

$$\therefore \quad \tilde{A} \equiv \underline{\underline{M}}^{-1} A^T \underline{\underline{M}}^T$$

We usually choose $\underline{\underline{M}}$ so that (v, u) is positive definite, so

$$\underline{\underline{M}}^T = M$$

$$\rightarrow \quad \underline{\underline{\tilde{A}}} = \underline{\underline{M}}^{-1} A^T \underline{\underline{M}}$$

SPACES

Consider the linear system:

$$y_1 = a_{11} x_1 + a_{12} x_2 + a_{13} x_3$$

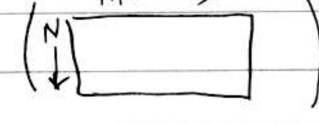
$$y_2 = a_{21} x_1 + a_{22} x_2 + A_{23} x_3$$

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \rightarrow \text{3-space}$$

$$\underline{y} = \underset{(2 \times 3)}{\underline{\underline{A}}} \quad \underline{x}$$

So what A does is to transform from 3-space to 2-space.

Consider the more general case: $N \times M$ rectangular matrix $\underline{\underline{A}}$ where $N < M$

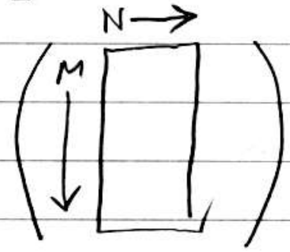


So $\underline{\underline{A}}$ operates on the set of vectors $\underline{u}$ of length M. and yields a set of vectors $\underline{w}$ of length N

$$\underset{\substack{(N \times 1) \\ N\text{-space}}}{\underline{w}} = \underset{(N \times M)}{\underline{\underline{A}}} \; \underset{\substack{(M \times 1) \\ M\text{-space}}}{\underline{u}}$$

Therefore $\underline{\underline{A}}$ maps from M-space to N-space.

- The adjoint $\underline{\underline{A}}^T$ is a $(M \times N)$ matrix



$\underline{\underline{A}}^T$ operates on vectors $\underline{v}$ of length N and yields vector $\underline{z}$ of length M

$$\underset{(M \times 1)}{\underline{z}} = \underset{(M \times N)}{\underline{\underline{A}}^T} \, \underset{(N \times 1)}{\underline{v}}$$

Therefore, $\underline{\underline{A}}^T$ map from N-space to M-space

- Suppose we want to solve:

$$\underline{y} = \underline{\underline{A}} \, x$$

given $\underline{\underline{A}}$ and $\underline{y}$. Here, $\underline{\underline{A}}$ $(N \times M)$ where $N < M$. That is, we have N equations for M unknowns.

Is there a solution? Yes, there is a "natural" solution related to adjoint of $\underline{\underline{A}}$

We look for solutions

$$\underset{\text{M-space}}{\underline{x}} = \underline{\underline{A}}^T \underset{\text{N-space}}{s}$$

Then, substituting

$$\underset{(N \times 1)}{\underline{y}} = \underset{(N \times N)}{\underset{(N \times M)(M \times N)}{\underline{\underline{A}} \, \underline{\underline{A}}^T}} \, \underset{(N \times 1)}{\underline{s}}$$

A familiar example: Consider the vertical component of vorticity

vorticity: $\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$

suppose we have a field of vorticity ζ or a vector $\underline{\zeta}$.
find the velocity?

Let look for the natural solution:

$$\zeta = \left(-\frac{\partial}{\partial y}, \frac{\partial}{\partial x}\right)\begin{pmatrix} u \\ v \end{pmatrix} = A\begin{pmatrix} u \\ v \end{pmatrix}$$

Solutions of the form $\begin{pmatrix} u \\ v \end{pmatrix} = A^T s$

$$A^T = \begin{pmatrix} \frac{\partial}{\partial y} \\ -\frac{\partial}{\partial x} \end{pmatrix}$$

so $\zeta = AA^T s = \left(-\frac{\partial}{\partial y} \; \frac{\partial}{\partial x}\right)\begin{pmatrix} \frac{\partial}{\partial y} \\ -\frac{\partial}{\partial x} \end{pmatrix} s$

$$= \left(-\frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial x^2}\right) s$$

$$= -\nabla^2 s$$

Let's identify $\psi = -s$, streamfunction

$\Rightarrow \zeta = \nabla^2 \psi$ where $u = -\frac{\partial \psi}{\partial y}$

$v = \frac{\partial \psi}{\partial x}$

- Recall that the $(N \times M)$ matrix $\underline{\underline{A}}$ $(N<M)$ has an operator equivalent in continuous space.

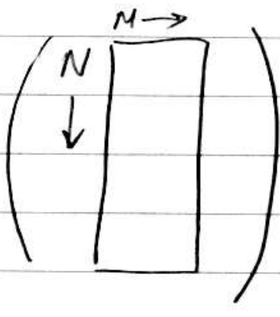
  → The operator A acts only on part of the space we say that only some dimensions of the space are "activated" by A.

- The remaining "non-activated" dimensions ⇒ "Null space"

- The adjoint A^{+} indentifies the activated space and ignores the null space.

<u>Overdetermined systems and "least-squares"</u>

* Consider the case where $\underline{\underline{A}}$ is an $(N \times M)$ matrix and $N>M$



$$y_1 = a_{11}x_1 + a_{12}x_2$$
$$y_2 = a_{21}x_1 + a_{22}x_2$$
$$y_3 = a_{31}x_1 + a_{32}x_2$$

$$\underline{Y} = \underline{\underline{A}}\,x$$

recall that the adjoint $\underline{\underline{A}}^T$ transforms from N-space into M-space, so we are tempting to solve

$$\boxed{\underline{\underline{A}}^T \underline{\underline{A}}\, x = \underline{\underline{A}}^T y}$$

To solve this problem we usually minimize the square of the differences.

$$J = (\underline{\underline{A}}\underline{x} - \underline{y})^T (\underline{\underline{A}}\underline{x} - \underline{y})$$

$$= \underline{x}^T \underline{\underline{A}}^T \underline{\underline{A}} \underline{x} - \underline{x}^T \underline{\underline{A}}^T \underline{y} - \underline{y}^T \underline{\underline{A}} \underline{x} + \underline{y}^T \underline{y}$$

at the extreme $\rightarrow$ $\frac{\partial J}{\partial \underline{x}} = 0$

$$\Rightarrow 2\underline{\underline{A}}^T \underline{\underline{A}} x - 2\underline{\underline{A}}^T \underline{y} = 0$$

$$\Rightarrow \boxed{\underline{\underline{A}}^T \underline{\underline{A}} \underline{x} = \underline{\underline{A}}^T \underline{y}}$$

Usefull rules of matrix calculus

$$\frac{d}{d\underline{x}}(\underline{\underline{A}}\underline{x}) = \underline{\underline{A}}^T$$

$$\frac{d}{d\underline{x}}(\underline{x}^T \underline{\underline{A}}) = A$$

$$\frac{d}{d\underline{x}}(\underline{x}^T \underline{x}) = 2x$$

$$\frac{d}{dx}(\underline{x}^T \underline{\underline{A}} x) = \underline{\underline{A}}\underline{x} + \underline{\underline{A}}^T x$$

Two-cases:

(1) $\underline{Y} = \underline{\underline{A}}\underline{x}$, $\underline{\underline{A}}$ $(N \times M)$ matrix, $N < M$

Natural solution given by solution of:

$$\underline{Y} = \underline{\underline{A}}\underline{\underline{A}}^T \underline{s}, \quad \underline{x} = \underline{\underline{A}}^T s$$

$\rightarrow$ underdetermined system, fewer observations than degrees of freedom

$\rightarrow$ Data assimilation in observation space. (representer method)

(2) $\underline{Y} = \underline{\underline{A}}\,\underline{x}$, $\underline{\underline{A}}$ is $(N \times M)$ matrix, $N > M$

least-squares solutions given by: $\underline{\underline{A}}^T \underline{y} = \underline{\underline{A}}^T \underline{\underline{A}}\,\underline{x}$

→ over determined system.
→ data assimilation in state-space (NWP).
4DVAR

Tutorial #2 Adjoint sensitivity Analysis

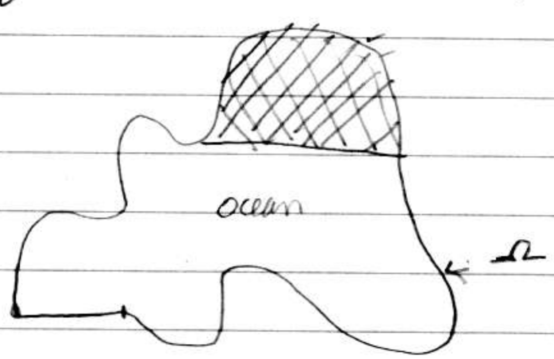
- Consider the state vector, $\underline{S}_0(t) = \begin{pmatrix} u \\ v \\ T \\ S \\ \varsigma \\ \vdots \end{pmatrix}$

which is a solution of the nonlinear equation

$$\frac{d\underline{S}_0}{dt} = N(\underline{S}_0) + \underline{f}_0(t)$$

↑ dynamical operator ↑ surface forcing

Subject to initial conditions: $\underline{S}_0 = (0)$
boundary conditions: $\underline{S}_{0\Omega}(t)$ Ω: physical boundary (close, open)



- Consider a differentiable, scalar function:

$$J(t) = G\left(\underline{s}_0(t)\right)$$

- We are interested on the sensitivity of J to variations in all possible values of $\underline{s}_0(0)$, $\underline{s}_{0\Omega}(t)$, and $\underline{f}_0(t)$.

- Consider "small" perturbations

$$\underline{s}(0) = \underline{s}_0(0) + \delta \underline{s}(0)$$
$$\underline{s}_\Omega(t) = \underline{s}_{0\Omega}(t) + \delta \underline{s}_\Omega(t)$$
$$\underline{f}(t) = \underline{f}_0(t) + \delta \underline{f}(t)$$

so the new solution $\underline{s}(t) = \underline{s}_0(t) + \delta \underline{s}(t)$ given by

$$\frac{d}{dt}\left(\underline{s}_0(t) + \delta \underline{s}(t)\right) = \underline{N}(\underline{s}_0 + \delta \underline{s}) + \underline{f}_0(t) + \delta \underline{f}(t)$$

A first-order Taylor expansion for

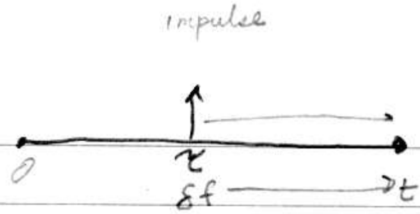
$\|\delta \underline{s}\| << \|\underline{s}_0\|$ yields

$$\frac{d\,\underline{s}_0(t)}{dt} + \frac{d}{dt}\delta \underline{s}(t) \simeq \underline{N}(\underline{s}_0) + \left.\frac{\partial \underline{N}}{\partial \underline{s}}\right|_{\underline{s}_0(t)} \delta \underline{s} + \underline{f}_0(t) + \delta \underline{f}(t) + \text{h.o.t}$$

The tangent linear equation:

$$\frac{d}{dt}\delta \underline{s} = \left.\frac{\partial \underline{N}}{\partial \underline{s}}\right|_{\underline{s}_0(t)} + \delta \underline{f}(t)$$

$$= \underline{\underline{A}}(t)\,\delta \underline{s} + \delta \underline{f}(t)$$



with solutions

$$\delta s(t) = R(0,t)\,\delta \underline{s}(0) + \int_0^t \underline{\underline{R}}(\tau,t)\,\underline{\underline{P}}\,\delta f(\tau)\,d\tau$$

↳ imposes the boundary conditions

subject to $\delta \underline{s}(0)$ and $\delta s_n(t)$.

Here $R(0,t)$ is the propagator matrix.

- Consider now the change in the function J:

$$J(t) + \delta J(t) = G\left(s_0(t) + \delta \underline{s}(t)\right)$$

$$\simeq G\left(S_0(t)\right) + \left(\frac{dG}{dS(t)}\right)^T\Bigg|_{S_0(t)} \delta \underline{s}(t)$$

$$\therefore\ \delta J(t) = \left(\frac{dG}{dS(t)}\Bigg|_{S_0(t)}\right)^T \delta \underline{s}(t)$$

$$= \delta \underline{s}^T(t)\left(\frac{dG}{d\underline{S}(t)}\Bigg|_{S_0(t)}\right)$$

$$= \delta \underline{s}^T(0)\,\underline{\underline{R}}^T(t,0)\,\frac{dG}{dS}\Bigg|_{S_0(t)} + \int_0^t \delta f^T(\tau)\,\underline{\underline{P}}^T\,\underline{\underline{R}}^T(t,\tau)\,\frac{dG}{dS}\Bigg|_{S_0(t)} d\tau$$

Consider the limit

$$\lim_{\delta \underline{s}(0)\to 0} \frac{\delta J(t)}{\delta \underline{S}(0)} = \frac{\partial J}{\partial \underline{S}(0)} = \underline{\underline{R}}^T(t,0)\,\frac{dG}{dS}\Bigg|_{S_0(t)}$$

Here $\underline{\underline{R}}^T(t,0)$ is the propagator of the corresponding adjoint.

To proceed further, we will discretize in time the equation for δJ

$$\delta J(N\Delta t) = \delta \underline{s}(0)\, \underline{\underline{R}}^T(N\Delta T, 0) \left.\frac{dG}{d\underline{s}}\right|_{\underline{s}_0(N\Delta t)}$$

$$+ \Delta t \sum_{i=0}^{N} \delta \underline{f}^T(i\Delta t)\, \underline{\underline{P}}^T \underline{\underline{R}}^T(N\Delta T, (N-i)\Delta t) \left.\frac{dG}{dS}\right|_{\underline{s}_0(N\Delta t)}$$

- Consider

$$\lim_{\delta f(i\Delta t)\to 0} \frac{\delta J(N\Delta t)}{\delta f(i\Delta t)} = \frac{\partial J}{\partial f(i\Delta t)} = \Delta t\, \underline{\underline{P}}^T \underline{\underline{R}}^T(N\Delta t, (N-i)\Delta t) \left.\frac{dG}{ds}\right|_{\underline{s}_0(N\Delta t)}$$

Choice of the Functional form of J

(1) Integrals over space: V: volume

$$J(t) = \frac{1}{V}\int_V s_0^2(t)\, dV$$

In discrete space:

$$J(t) = \underline{s}_0^T(t)\, \underline{\underline{X}}\, \underline{s}_0(t)$$

where
$$\underline{\underline{X}} = \begin{pmatrix} \frac{\Delta V_1}{V} & & 0 \\ & \frac{\Delta V_2}{V} & \\ 0 & & \ddots \end{pmatrix}$$

- If $\underline{s}(t) = \underline{s}_0(t) + \delta \underline{s}(t)$

$$\Rightarrow J(t) + \delta J(t) = \left(\underline{s}_0(t) + \delta \underline{s}(t)\right)^T \underline{\underline{X}} \left(\underline{s}_0(t) + \delta s(t)\right)$$

$$\Rightarrow \delta J(t) = 2\, \delta \underline{s}^T(t)\, \underline{\underline{X}}\, \underline{s}_0(t)$$

Then

- $$\frac{\partial J}{\partial \underline{s}(0)} = 2 \underline{\underline{R}}^T(t,0) \underline{\underline{X}} \underline{S}_0'(t)$$

- $$\frac{\partial J}{\partial \underline{f}(\tau)} = 2\Delta t \, \underline{\underline{P}}^T \underline{\underline{R}}^T(t,\tau) \underline{\underline{X}} \underline{S}_0'(t)$$

(2) Integrals over space and time:

Suppose J is of the form

$$J = \frac{1}{VT} \int_0^T \int_V S_0'^2(t) \, dv \, dt$$

in discrete form

$$J \simeq \frac{\Delta t}{K\Delta t} \sum_{k=0}^{K} \underline{S}_0'^T(k\Delta t) \underline{\underline{X}} \, \underline{S}_0(k\Delta t) \qquad T = K\Delta t$$

- $S(t) = S_0(t) + \delta \underline{s}(t)$

$$\frac{\partial J}{\partial \underline{S}(0)} = \frac{2}{K} \underbrace{\sum_{k=0}^{K} \underline{\underline{R}}^T(0,k\Delta t) \underline{\underline{X}} \, \underline{S}_0(k\Delta t)}$$

solution of the adjoint equations forced $\dfrac{2 \underline{\underline{X}} S_0(k\Delta t)}{K \Delta t}$.

$$\frac{dJ}{d\underline{f}(j\Delta t)} = \frac{2\Delta t}{K} \underbrace{\sum_{k=j}^{K} \underline{\underline{P}}^T \underline{\underline{R}}^T(k\Delta t, (k-j)\Delta t) \underline{\underline{X}} \, S_0'(k\Delta t)}$$

($j\Delta t$ ↑ τ)

Solution of the adjoint forced by

$$\frac{2 \underline{\underline{P}}^T \underline{\underline{X}} \, S_0'(k\Delta t)}{K}$$

Suppose that we write the tangent linear equation as

$$\frac{d\underline{s}}{dt} = \underline{\underline{A}}\,\underline{s} + \underline{f}(t)$$

integrating factor of $e^{-\underline{\underline{A}}t}$

$$\rightarrow \quad \underline{s}(t) = e^{\underline{\underline{A}}t}\underline{s}(0) + \int_0^t e^{\underline{\underline{A}}(t-\tau)}\underline{f}(\tau)\,d\tau$$

The adjoint equations

$$-\frac{d\underline{s}^+}{dt} = \underline{\underline{A}}^+\underline{s}^+ + \underline{f}^+(t)$$

$$\rightarrow \quad \underline{s}^+(0) = e^{\underline{\underline{A}}^+t}\underline{s}^+ + \int_0^t e^{\underline{\underline{A}}^+t} f^+(\tau)\,d\tau$$

$A^+ = A^T$ (L-2 norm)

$\underline{R} = e^{\underline{A}t}$

$\underline{\underline{R}}^T = (e^{\underline{\underline{A}}t})^T = e^{\underline{\underline{A}}^T t}$

Tutorial # 3

"Data Assimilation"

A glorified Least-squares minimization problem or minimum variance estimate, or maximum likehood estimate.

- Two methods: (i) model-space search
  (ii) observation-space search

- Two flavors (i) strong constraint (no model errors)
  (ii) weak constraint

Least-square Minimization Example:
(minimum variance, maximum likelihood)

- Suppose we have two independent observations of x, x_1 and x_2

- Form an estimate, $\boxed{x_a = a_1 x_1 + a_2 x_2}$

- Each observation will have associated errors:

$$\varepsilon_1 = x_1 - x_t \qquad \varepsilon_2 = x_2 - x_t \qquad x_t : x \text{ true}$$

- assumptions, Gaussian, random errors

(1) $\langle \varepsilon_1 \rangle = \langle \varepsilon_2 \rangle = 0$

(2) Assume an unbiased, Gaussian estimate:

$$\langle \varepsilon_a \rangle = \langle x_a - x_t \rangle = 0$$

$$\Rightarrow \quad a_1 + a_2 = 1$$

(3) No correlation between measurement errors

$$\langle \varepsilon_1 \varepsilon_2 \rangle = 0$$

$$\Rightarrow \quad \langle \varepsilon_a^2 \rangle = a_1^2 \langle \varepsilon_1^2 \rangle + a_2^2 \langle \varepsilon_2^2 \rangle$$

$$\boxed{\sigma_a^2 = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2}$$

σ: standard deviation of error.

<u>Problem</u>: find a_1 and a_2 that minimizes σ_a^2 but subject to the condition that $a_1 + a_2 = 1$

Constrained minimization problem can be solved using method of Lagrange Multipliers

$$\mathcal{L} = \sigma_a^2 + \lambda (a_1 + a_2 - 1)$$

λ: unknown Lagrange multiplier

at the extrema we have:

$$\left.\begin{aligned} \frac{\partial \mathcal{L}}{\partial a_1} = 0 \qquad & 2 a_1 \sigma_1^2 + \lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial a_2} = 0 \qquad & 2 a_2 \sigma_2^2 + \lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} = 0 \qquad & a_1 + a_2 - 1 = 0 \end{aligned}\right\}$$

$$\boxed{x_a = \frac{\sigma_1^{-2}}{(\sigma_1^{-2} + \sigma_2^{-2})} x_1 + \frac{\sigma_2^{-2}}{(\sigma_1^{-2} + \sigma_2^{-2})} x_2}$$

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

Generalizing for M-observations, x_n with error estimates G_n^2 from $n=1,\ldots,M$

$$x_a = \frac{\sum_{n=1}^{M} G_n^{-2} x_n}{\sum_{n=1}^{M} G_n^{-2}} \qquad G_a^2 = \left[\sum_{n=1}^{M} G_n^{-2}\right]^{-1}$$

4-Dimentional Variational Data Assimilation

Notation:

(i) $\underset{\sim}{S} = (u, v, T, S, \zeta, \ldots)^T$ roms state-vector

(ii) NL ROMS: $\frac{d\underset{\sim}{S}}{dt} = N^P(\underset{\sim}{S}) + \underline{f}(t)$

subject $\underset{\sim}{S}(0)$, $\underset{\sim}{S}_a(t)$

(iii) Observations: $\underline{Y}_i$ at time t_i with observation error covariance $\underline{\underline{O}}$

(iv) Model equivalent at observation point

$$\underline{\underline{H}}_i \underset{\sim}{S}(t_i) = \underline{\underline{H}}_i S_i$$

(v) Background, S_b with associated error covariance $\underline{\underline{B}}$.
(typically $\underset{\sim}{S}_b$ will be from a model solution)

(vi) Desired estimate, $\underset{\sim}{S}_a$

Seek to minimize: Cost function

$$J(\underline{S}) = \frac{1}{2}\left(\underline{S}(0) - \underline{S}_b\right)^T \underline{\underline{B}}^{-1}\left(\underline{S}(0) - \underline{S}_b\right)$$

$$+ \frac{1}{2}\sum_{i=1}^{M}\left(\underline{\underline{H}}_i \underline{S}_{i_a} - \underline{y}_i\right)^T \underline{\underline{O}}^{-1}\left(\underline{\underline{H}}_i \underline{S}_{i_a} - \underline{y}_i\right)$$

<u>Constraints:</u>

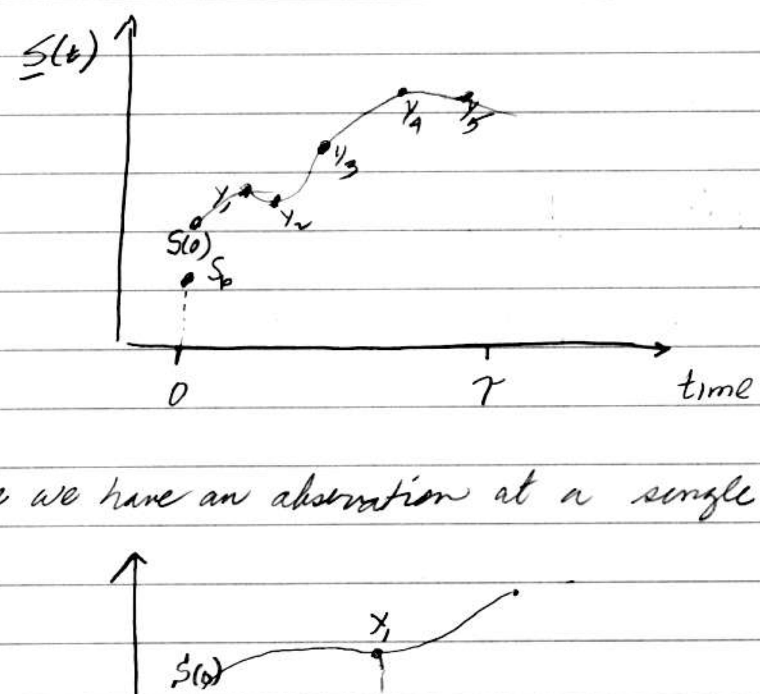
- $\underline{S}$ satisfies NLRDMS <u>exactly</u> (strong constraint)

<u>Problem:</u>

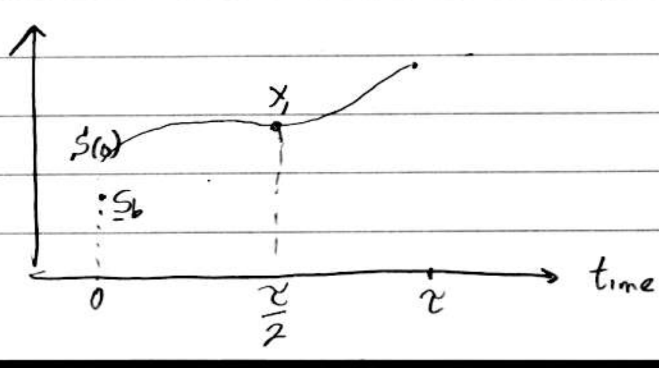
- Find $\underline{S}(t)$ that minimizes $J(\underline{S})$ and satisfies

$$\frac{d\underline{S}}{dt} - \mathcal{N}(\underline{S}) - \underline{f}(t) = 0,$$

subject to $\underline{S}(0)$ and $\underline{S}_a(t)$



assume the we have an observation at a single time



To Solve the problem:

$$\mathcal{L} = J(\underline{s}) + \sum_{i=0}^{N} \lambda_i^T \left(\frac{d\underline{s}_i}{dt} - N(\underline{s}_i) - \underline{f}_i \right)$$

where $t = [0, \tau] \equiv [0, N\Delta T]$

$\underline{\lambda}_i \equiv \underline{\lambda}(i\Delta t)$
$\underline{f}_i \equiv \underline{f}(i\Delta t)$
...

- Suppose we change $\underline{s}(0)$ by $\delta \underline{s}(0)$: $\rightarrow \delta \underline{s}(t)$
  consider variations $\delta\lambda(t)$ in $\lambda(t)$

The resulting first variation $\delta\mathcal{L}$: a single observation at $\frac{\tau}{2}$

$$\delta\mathcal{L} \simeq \delta\underline{s}^T(0)\,\underline{\underline{B}}^{-1}\left(\underline{s}(0) - \underline{s}_b\right) + \delta\underline{s}^T\left(\tfrac{\tau}{2}\right)\underline{\underline{H}}^T\underline{\underline{O}}^{-1}\left(\underline{\underline{H}}\,\underline{s}\left(\tfrac{\tau}{2}\right) - y\right)$$

$$+ \sum_{i=0}^{N} \underline{\lambda}_i^T \left(\frac{d\delta s_i}{dt} - \left(\frac{\partial N}{\partial s}\right)\delta s \right)$$

$$+ \sum_{i=0}^{N} \delta\underline{\lambda}_i^T \left(\frac{d s_i}{dt} - N(s_i) - f_i \right)$$

Green's identity →

This can be written as

$$\underbrace{\sum_{i=0}^{N} \frac{d}{dt}\left(\underline{\lambda}_i^T \delta\underline{s}_i\right)}_{\delta s^T(\tau)\lambda(\tau) - \delta s^T(0)\underline{\lambda}(0)} + \underbrace{\sum_{i=0}^{N} \delta\underline{s}_i^T \left(-\frac{d\lambda_i}{dt} - \left(\frac{\partial N}{\partial s}\right)^T \underline{\lambda}_i \right)}_{\text{adjoint equation}}$$

at the extrema of $\mathcal{L}$, we require

$$\frac{\partial \mathcal{L}}{\partial \underline{S}_i} = 0 \qquad \lim_{\delta S_i \to 0} \frac{\delta \mathcal{L}}{\delta S_i}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = 0 \qquad \lim_{\delta \lambda_i \to 0} \frac{\delta \mathcal{L}}{\delta \lambda_i}$$

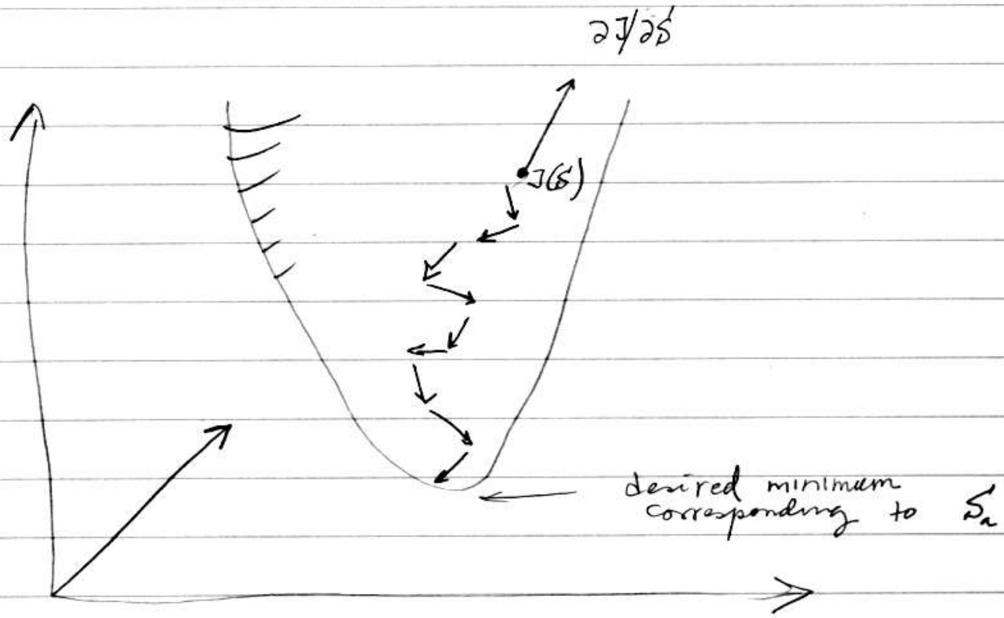
we get four equations:

(1) $\frac{d \underline{S}_i'}{dt} - N(\underline{S}_i) - f_i = 0$ NLROMS

(2) $-\frac{d\lambda_i}{dt} - \left(\frac{\partial N}{\partial \underline{S}'}\right)^T \lambda_i + \delta_{i,\frac{N}{2}} \underline{\underline{H}}^T \underline{\underline{O}}^{-1} \left(\underline{\underline{H}}\, \underline{S}'_{\frac{N}{2}} - \underline{Y}\right) = 0$ ADROMS

(3) $\underline{\lambda}(\tau) = 0$ the initial condition of the ADM at $t=\tau$ is zero.

(4) $\underline{\underline{B}}^{-1}\left(\underline{S}(0) - \underline{S}_b\right) - \underline{\lambda}(0) = \frac{\partial J}{\partial \underline{S}(0)} = 0$ $\left(\leftarrow \frac{\partial \mathcal{L}}{\partial S(0)}\right)$



NWP

$$\underset{\sim}{S}_a(t_0) \xrightarrow{6\text{ hours}} S_b \equiv S(t_b + 6\text{ hours})$$

$$J = \frac{1}{2}\left(\underline{S}(0) - \underline{S}_b\right)^T \underline{\underline{B}}^{-1}\left(\underline{S}(0) - \underline{S}_b\right)$$

$$+ \frac{1}{2}\sum_{i=1}^{M}\left(\underline{\underline{H}}_i \underline{S}_i - y_i\right)^T \underline{\underline{O}}^{-1}\left(\underline{\underline{H}}_i \underline{S}_i - y_i\right)$$

Assume

Incremental 4-dimensional variational data assimilation

ISADVAR

$$\underline{S}_i = \underline{S}_{bi} + \delta \underline{S}_i$$

Therefore,

$$J \simeq \frac{1}{2}\,\delta\underline{S}(0)^T \underline{\underline{B}}^{-1} \delta\underline{S}(0) + \frac{1}{2}\sum_{i=1}^{N}\left(\underline{\underline{G}}_i\,\delta\underline{S}(0) - \underline{d}_i\right)^T \underline{\underline{O}}^{-1}$$

$$\left(\underline{\underline{G}}_i\,\delta\underline{S}(0) - \underline{d}_i\right)$$

$$\underline{\underline{G}}_i = \underline{\underline{H}}_i\ \underline{\underline{R}}(0, t_i)$$

TLROMS

$$\underline{d}_i = \underline{y}_i - \underline{\underline{H}}_i\,\underline{S}_{bi}$$

Basic Algorithm for ISADVAR

(1) Choose Initial guess, $\underset{\sim}{S}(0) = \underset{\sim}{S}_b$

(2) Integrate NLROMS, "J", $\underset{\sim}{S}_i(t)$, $t = [0, \tau]$

(a) Choose $\delta \underset{\sim}{S}(0) = 0$

(b) Integrate TLROMS, J, $t = [0, \tau]$

(c) Integrate ADROMS forced by the adjoint misfit $\rightarrow \frac{\partial J}{\partial \underset{\sim}{S}(0)}$

(d) Use a conjugate gradient algorith to determine the the downgradient directions to $\delta \underset{\sim}{S}(0)$

conjugate gradient

(i) Which way? Preconditioning

(ii) How far? Line search

Introduce a new variable

$$\underline{v} = B^{-1/2} \delta s \qquad \text{so} \qquad J_b = \frac{1}{2} \underline{v}^T(0)\, \underline{v}(0)$$

$$\delta s = \underline{\underline{B}}^{1/2}\, \underline{v}$$

Summary

- Model: $\frac{d\underline{s}}{dt} = N(\underline{s}) + f(t)$, $\underbrace{\underline{s}(0),\ \underline{s}_\Omega(t),\ f(t)}_{\text{control variables}}$

  NLROMS

- Background: $\underline{s}_b$
- Observations: y_i
- Cost function:

$$J = \frac{1}{2}\left(\underline{s}(0) - \underline{s}_b\right)^T \underline{\underline{B}} \left(\underline{s}(0) - \underline{s}_b\right) +$$

$$\frac{1}{2}\sum_i \left(\underline{\underline{H}}_i \underline{s}_i - y_i\right)^T \underline{\underline{O}}^{-1} \left(\underline{\underline{H}}_i \underline{s}_i - y_i\right)$$

- If $\underline{s}_b$ is a prior model forecast, then we can assume

$$\underline{s} = \underline{s}_b + \delta\underline{s}$$

- Assuming (hoping!) $\|\delta\underline{s}\| \ll \|\underline{s}_b\|$, apply tangent linear assumption

$$\frac{d\delta\underline{s}}{dt} = \left.\frac{\partial N}{\partial \underline{s}}\right|_{\underline{s}_b} \Rightarrow \quad \delta\underline{s}(t) = \underline{\underline{R}}(0,t)\,\delta\underline{s}(0)$$

TLROMS

- Cost function becomes

$$J \simeq \underbrace{\frac{1}{2}\delta\underline{s}(0)\,\underline{\underline{B}}^{-1}\,\delta s(0)}_{J_b} + \underbrace{\frac{1}{2}\sum_i \left(\underline{\underline{H}}_i \underline{\underline{R}}\,\delta\underline{s}(0) - \underline{d}_i\right)^T \underline{\underline{O}}^{-1} \left(\underline{\underline{H}}_i \underline{\underline{R}}\,\delta s(0) - \underline{d}_i\right)}_{J_o}$$

where $\underline{d}_i = y_i - H_i \underline{s}_{b_i}$

- Cost function gradient

$$\frac{\partial J}{\partial \underline{s}(0)} = \frac{\partial J_b}{\partial \underline{s}(0)} + \frac{\partial J_o}{\partial \underline{s}(0)}$$

— Given by the solution of the forced adjoint model (ADROMS)

- Find $\delta s(0)$ that minimizes J iteratively

→ preconditioning is a major issue

- Argument: J_b largely dictates the structure of J so precondition using $\underline{\underline{B}}$ via a change of variable

$$\underline{V} = \underline{\underline{B}}^{-1/2}\,\delta\underline{s}$$

$$\Rightarrow \quad J = \frac{1}{2}\,\underline{V}^T(0)\,V(0) + \frac{1}{2}\sum_i \left(\underline{\underline{H}}_i \underline{\underline{R}}\, \underline{\underline{B}}^{1/2}\underline{V}(0) - \underline{d}_i\right)^T O^{-1}\left(\underline{\underline{H}}_i \underline{\underline{R}}\, \underline{\underline{B}}^{1/2}\underline{V}(0) - \underline{d}_i\right)$$

- $\underline{\underline{B}}$ is very large and unwieldy!

- Procedurally $\underline{\underline{B}}$ is factorized

$$\underline{\underline{B}} = \underline{\underline{K}}_b\,\underline{\underline{B}}_u\,\underline{\underline{K}}_b^T$$

$\underline{\underline{K}}_b$ — balanced operator; $\underline{\underline{B}}_u$ — univariate background covariance

$$\underline{\underline{B}}_u = \underline{\underline{\Sigma}}\,\underline{\underline{C}}\,\underline{\underline{\Sigma}}$$

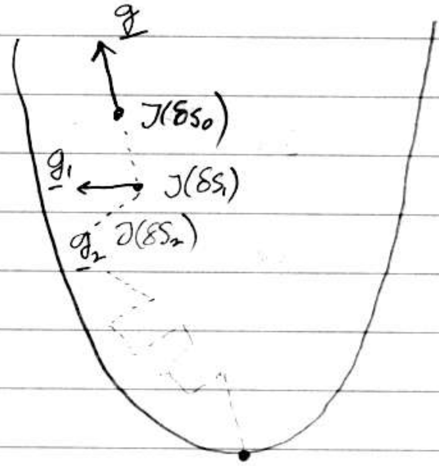
$\underline{\underline{\Sigma}}$ — diagonal matrix standard deviation; $\underline{\underline{C}}$ — correlation matrix

Factorization of $\underline{\underline{C}} = \underline{\underline{C}}^{1/2}\,C^{T/2}$

Separable $\underline{\underline{C}}^{1/2} = \underline{\underline{C}}_v^{1/2}\,\underline{\underline{C}}_h^{1/2}$

$\underline{\underline{C}}_v^{1/2}$ — vertical correlation; $\underline{\underline{C}}_h^{1/2}$ — horizontal correlation

$\underline{\underline{c}}_v$, $\underline{\underline{c}}_h$ are modelled as solution of approximate diffusion



every δs is actually a ✓ because where are in minimization space.

Conjugate Gradient method:

$\delta \underline{s}_0(0) = \underline{a}$ a specified starting point

$\underline{g}_0 \equiv \partial J / \partial \underline{s}_0(0)$ $\underline{d}_0 = -\underline{g}_0$ starting value of the gradient and conjugate direction.

$\delta \underline{s}_1(0) = \delta \underline{s}_0(0) - \alpha_1 \underline{g}_0$ α: step size

$\underline{g}_1 \equiv \partial J / \partial \underline{s}_1(0)$ $\underline{d}_1 = -\underline{g}_1 + \beta_0 \underline{d}_0$, $\beta_0 = g_1^T g_1 / g_0^T g_0$

$\delta \underline{s}_2(0) = \delta \underline{s}_1(0) + \alpha_2 \underline{d}_1$

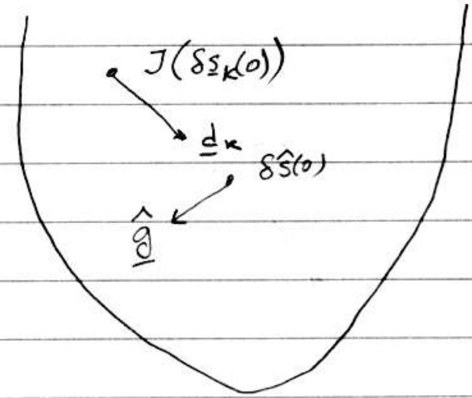
$\underline{g}_2 \equiv \partial J / \partial \underline{s}_2(0)$ $\underline{d}_2 = -\underline{g}_2 + \beta_1 \underline{d}_1$ $\beta_1 = g_2^T g_2 / g_1^T g_1$

⋮

$\delta \underline{s}_n(0) = \delta \underline{s}_{n-1}(0) + \alpha_n \underline{d}_{n-1}$

$\underline{g}_n = \partial J / \partial \underline{s}_n(0)$ $\underline{d}_n = -\underline{g}_n + \beta_n \underline{d}_{n-1}$, $\beta_n = g_n^T g_n / g_{n-1}^T g_{n-1}$

CONGRAD



Since the cost function is quadratic the gradient can be recomputed.

$$J = \frac{1}{2} a x^2 + b x + c$$

$$\frac{dJ}{dx} = ax + b$$

$$\frac{d^2 J}{dx^2} = a$$

$$\delta \hat{\underline{s}}(0) = \delta \underline{s}_K(0) + \tau \underline{d}_K$$

τ : trial step size

The gradient along the direction $\underline{d}_K$, $\hat{g}_K$ is compute by runing the adjoint model forced by $\delta \hat{\underline{s}}$

$$\tilde{\underline{g}} = \underline{g}_K + \tilde{\alpha} \frac{(\hat{g}_K - g_K)}{\tau}$$

orthogonalize with all previous gradients.

- Find $\tilde{\alpha}$ for which $\underline{d}_K^T \tilde{\underline{g}} = 0$

$$\therefore \quad \underline{d}_K^T \left[g_K + \frac{\tilde{\alpha}}{\tau} (\hat{g}_K - g_K) \right] = 0$$

$$\Rightarrow \quad \tilde{\alpha} = \frac{\tau \underline{d}_K^T g_K}{\underline{d}_K^T (g_K - \hat{g}_K)}$$: optimum step size

Incremented formulation

$$J = \frac{1}{2}\left(\underline{S}(0) - \underline{S}_b\right)^T \underline{\underline{B}}^{-1}\left(\underline{S}(0) - \underline{S}_b\right)$$

$$+ \frac{1}{2}\sum_i \left(\underline{\underline{H}}_i \underline{\underline{R}}\, \delta \underline{S}(0) - \underline{d}_i\right)^T \underline{\underline{O}}^{-1}\left(\underline{\underline{H}}_i \underline{\underline{R}}\, \delta \underline{S}(0) - \underline{d}_i\right)$$

- Let $\underline{V} = \underline{\underline{B}}^{1/2}\, \delta \underline{S}(0) \Rightarrow \delta \underline{S}(0) = \underline{\underline{B}}^{1/2}\underline{V}$

- Therefore

$$J = \frac{1}{2}\underline{V}^T(0)\underline{V}(0) + \frac{1}{2}\sum_i \left(\underline{\underline{H}}_i \underline{\underline{R}}\, \underline{\underline{B}}^{1/2}\underline{V} - \underline{d}_i\right)^T \underline{\underline{O}}^{-1}\left(\underline{\underline{H}}_i \underline{\underline{R}}\, \underline{\underline{B}}^{1/2}\underline{V}(0) - \underline{d}_i\right)$$

$$\frac{\partial^2 J}{\partial V^2(0)} = \underline{\underline{I}} + \sum_i \underline{\underline{B}}^{1/2}\underline{\underline{R}}^T \underline{\underline{H}}^T \underline{\underline{O}}^{-1}\underline{\underline{H}}_i \underline{\underline{R}}\, \underline{\underline{B}}^{1/2}$$

$$= \mathcal{H} \quad \text{Hessian matrix}$$

Let

$$\underline{\underline{\mathcal{H}}} = \underline{\underline{E}}\, \underline{\underline{\Lambda}}\, E^T$$

$\underline{\underline{E}} = (e_i)$: matrix of eigenvectors of $\underline{\underline{\mathcal{H}}}$

$\underline{\underline{\Lambda}} = \mathrm{diag}(\lambda_i)$: diagonal matrix of eigenvectors

Suppose that we can calculate the leading eigenvectors/eigenvalues of $\mathcal{H}$

$$\underline{\underline{\mathcal{H}}}_K = \underline{\underline{E}}_K \underline{\underline{\Lambda}}_K \underline{\underline{E}}_K^T$$

K: number of eigenvectors

- More generally we can write

$$\underline{\underline{\mathcal{H}}} = \underline{\underline{Q}}\, \underline{\underline{T}}\, \underline{\underline{Q}}^T$$

$\underline{\underline{Q}} = (\underline{q}_i)$: matrix of orthnormal vectors

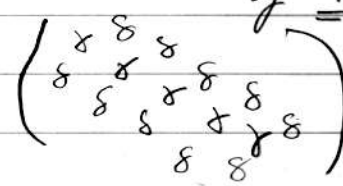
$\underline{\underline{T}}$: tridiagonal matrix

- There is a relationship between E and Q

$$T = Z \Lambda Z^T$$

Z = eigenvectors matrix of T

$$Q Z = E$$



- we can conveniently construct a set of q_i vectors during each iteration of the inner loop using the so called Lanczos method.

$$\mathcal{H}_k = Q_k T_k Q_k^T$$

$$\Rightarrow \quad \mathcal{H} Q_k = Q_k T_k + \gamma_k q_{k+1} \varepsilon_k^T \qquad \varepsilon_k^T = (00\cdots 1 \cdots 00) \quad k^{th}$$

- Lanczos recurrence:

$$\mathcal{H} q_{k+1} = \gamma_{k+1} q_{k+2} + \delta_{k+1} q_{k+1} + \gamma_k q_k$$

CG Method

$$\delta s_{k+1}(0) = \delta s_k(0) + \alpha_k d_k$$

$$d_{k+1} = - g_{k+1} + \beta_{k+1} d_k$$

Eliminate d_k:

$$\frac{1}{\alpha_{k+1}} \left(\delta s_{k+2} - \delta s_{k+1} \right) = - g_{k+1} + \left(\frac{\beta_{k+1}}{\alpha_k} \right) \left(\delta s_{k+1} - \delta s_k \right)$$

Multiply both sides by $\mathcal{H}$

$$\frac{1}{\alpha_{k+1}}\left(\underline{g}_{k+2} - \underline{g}_{k+1}\right) = -\underline{\underline{\mathcal{H}}}\,\underline{g}_{k+1} + \left(\frac{\beta_{k+1}}{\alpha_k}\right)\left(\underline{g}_{k+1} - \underline{g}_k\right)$$

let $$q_k = \frac{\underline{g}_k}{(\underline{g}_k^T g_k)}$$

yielding

$$\boxed{\underline{\underline{\mathcal{H}}}\, q_{k+1} = \gamma_{k+1}\, \underline{q}_{k+2} + \delta_{k+1}\, \underline{q}_{k+1} + \gamma_k\, q_k}$$

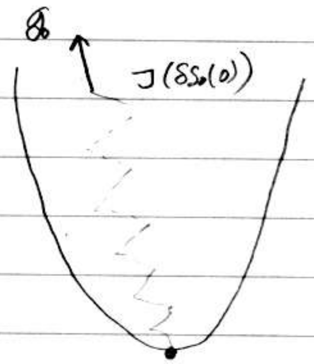
where $$\delta_{k+1} = \left(\frac{1}{\alpha_{k+1}} + \frac{\beta_{k+1}}{\alpha_k}\right)$$

$$\gamma_k = -\frac{\sqrt{\beta_{k+1}}}{\alpha_k}$$

$$V = \underline{V}_b + \underline{v}$$

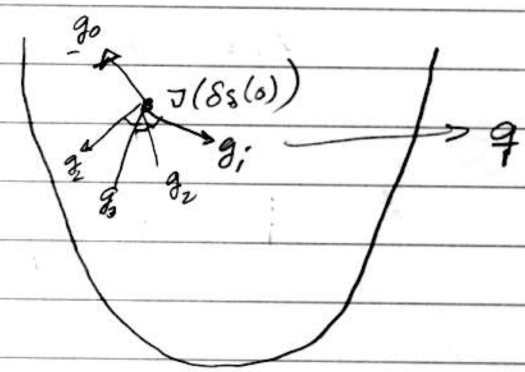
$$= V_b - \underline{\underline{Q}}_k\, T^{-1} Q^T g_0$$

old



CG approach

New



Lanczos

$J = ax^2 + bx + c$

$\frac{dJ}{\partial x} = ax + b$

$\frac{d^2 J}{\partial x^2} = a$

change of variables

$w = \underline{\underline{H}}_K^{-1} \underline{V}$

Basic ISADVAR Algorithm

① We need a background state from previous analysis; $S_b(0)$

② Run NLROMS and get $S_b(t)$ $\quad t = [0, \tau]$

ⓐ choose $\delta s(0) = 0$, $v(0) = 0$

→ convolution $\delta \underline{s}(0) = \underline{\underline{B}}^{1/2} v(0)$ convolution

ⓑ Run TLROMS and compute J $\quad t = [0, \tau]$

ⓒ Run ADROMS $t = [\tau, 0]$ to compute $h_0 = \frac{\partial J_0}{\partial \underline{S}(0)}$

convolve: $g = \underline{v} + B^{T/2} h_0$

ⓓ conjugate Gradient / Lanczos algorithm and give us a new value of $V(0)$

* Compute e's and approximate $\mathcal{H}_k$
* Preconditioning by $\underline{\underline{H}}_k^{-1}$ → w_k-space
* compute a new value of $\underline{v}(0)$.

(Inner Loop: ⓐ–ⓓ; Outer Loop: ②–③)

③ Compute new NLM initial conditions

$$\underline{S}(0) = \underline{S}_b - \underline{B}^{1/2} \left[\underline{\underline{Q}}_k \underline{\underline{T}}_k^{-1} \underline{\underline{Q}}_k^T g_0 \right]$$

Balance Operator

$$\underline{\delta x} = \underline{\underline{K}}\, \underline{\delta x_u}$$

$$\underline{\underline{B}} = \langle \underline{\delta x}\, \underline{\delta x}^T \rangle$$

$$= \langle \underline{\underline{K}}\, \underline{\delta x_u}\, \underline{\delta x}^T\, \underline{\underline{K}}^T \rangle$$

$$= \underline{\underline{K}} \langle \underline{\delta x_u}\, \underline{\delta x_u}^T \rangle \underline{\underline{K}}^T$$

$$= \underline{\underline{K}}\, B_u\, K^T$$

NL $\rightarrow$ $\underline{x}$

TL $\rightarrow$ $\underline{\delta x}$ $\longrightarrow$ $\underline{\underline{O}}^{-1}(\delta x - d)$

AD $\rightarrow$ $\partial J / \partial \underline{\delta x}$

Minimize in uncorrelated space

$$\underline{v} = \underline{\underline{B_u}}^{-1/2}\, \underline{\delta x_u}$$

$$\underline{\underline{B_u}}^{1/2} = \underline{\underline{C}}^{1/2}\, \underline{\underline{\Sigma_n}}^{1/2}$$

$$\underline{\delta x} = \underline{\underline{K}}\, \underline{\delta x_u} = \underline{\underline{K}}\, \underline{\underline{B_u}}^{1/2}\, \underline{v}$$

$$\frac{\partial J}{\partial \underline{v}} = \frac{\partial \delta x}{\partial \underline{v}} \frac{\partial J}{\partial \underline{\delta x}} = \underline{\underline{B_u}}^{T/2}\, \underline{\underline{K}}^T \frac{\partial J}{\partial \underline{\delta x}}$$

ISADVAR

outer: n = 1, Nouter

NL (x_b) → $\underline{x}$

inner = 1

$$\delta \underline{x}_u = \delta \underline{x} = \underline{v} = 0$$

TL → $\underline{\underline{O}}^{-1}(\delta \underline{x} - \underline{d})$

AD → $\partial J / \partial \delta \underline{x}$

$$\frac{\partial J}{\partial v} = \underbrace{\underline{\underline{B}}_u^{T/2} (\underline{\underline{K}}^T)^n}\, \partial J / \partial \delta \underline{x}$$

ad_balance
ad_convolution
ad_variability

$(K^T)^n$ it is outer loop dependent

CG → new $\underline{V}$

$$\delta \underline{x} = \underbrace{\underline{K}^{n+} B_u^{1/2}}\, \underline{V}$$

tl_variability
tl_convolution
tl_balance

K^n outer loop dependent

How to compute the standard deviation of the unbalance field

$$\Sigma_S = S - S_b$$

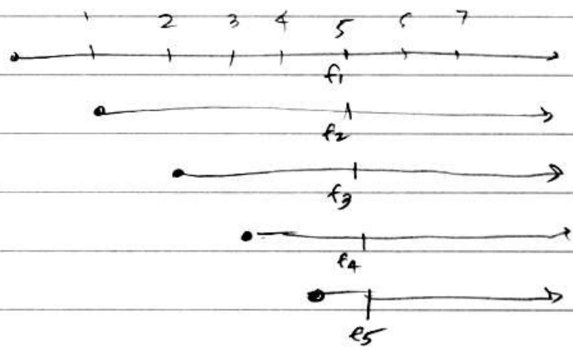
$$S = S_b + \Delta S$$

↓ Basic state

$$\delta S = \delta S_u + \delta S_B$$

$$\therefore \delta S_u = \delta S - \delta S_B$$

NMC method



$$\left.\begin{array}{l} f_5 - f_1 \\ f_5 - f_2 \\ f_5 - f_3 \\ \vdots \end{array}\right\} \delta S = \delta S_B + \delta S_u$$

Then compute Σ_{S_u} the unbalance standard deviation.